Final draft chapter for Pryor (Ed.), *Managing Research Data* (pp. 141-151). London: Facet.

# Roles and Responsibilities – Libraries, Librarians and Data

*Sheila Corrall, Information School, University of Sheffield*

> 'E-science has the potential to be transformational within research libraries by impacting their operations, functions, and possibly even their mission.'
>
> (Lougee et al., 2007, p. 3)

## Introduction

The management of the research data generated by e-science/e-research has replaced open access to scholarly publications as the hot topic on the academic library and information services agenda. National and international bodies have issued a succession of reports, policies and guidance on dealing with the 'data deluge' that have flagged the need for concerted action by research organisations. Government and other official publications have often identified roles for library and information professionals in managing data alongside the other information and knowledge resources that libraries manage or provide for their communities. Professional associations in the library world have responded positively to such suggestions and university librarians are now starting to develop services or get involved in projects to explore what libraries can do to support researchers in meeting new requirements of funding agencies to facilitate access to their data.

Managing research data continues to be an emergent area of activity where responsibilities and practices within libraries are generally not yet firmly established, especially in the United Kingdom. However, there have been significant developments within the past few years, particularly in the United States, as a result of nationally funded project work and new requirements of research funding agencies for data management plans to be submitted with grant applications. Many libraries have seized these opportunities to form new partnerships and develop new services, which has in turn generated a growing body of literature on the subject offering insights and pointers for practitioners in other countries. In this chapter, we review the opportunities offered and the challenges presented for libraries and librarians in the research data arena, with particular reference to published reports and case studies of emerging practice, supplemented by evidence from university and library websites.

Some commentators have questioned whether library staff have the knowledge and skills needed to fulfil the roles suggested, so we look specifically at connections between research data management and established library roles and responsibilities to explore whether research data management represents another incremental step in professional practice or a true paradigm shift in collection development and service delivery requiring fundamental rethinking of roles, responsibilities and competencies to create 'next-generation librarianship', drawing where possible on the experiences and opinions of practitioners already involved in the field. Finally, we discuss professional education and continuing development needs for library engagement with research data, again referring particularly to initiatives already taken in the US.

## The case for library engagement

Several expert commentators in the information field have argued that the problems faced by institutions in managing research data offer an important and attractive opportunity for libraries to redefine their role in supporting research and to develop closer relationships with their research community. Hey and Hey (2006, p. 526) suggest that if librarians can respond

effectively to the challenge, 'the e-Science revolution will put libraries and repositories center stage in the development of the next generation research infrastructure'. Swan and Brown (2008, p. 2) in their report on workforce development for the UK Joint Information Systems Committee similarly see data management as a strategic issue for libraries and librarians:

> 'The role of the library in data-intensive research is important and a strategic repositioning of the library with respect to research support is now appropriate. We see three main potential roles for the library: increasing data-awareness amongst researchers; providing archiving and preservation services for data within the institution through institutional repositories; and developing a new professional strand of practice in the form of data librarianship.'

Other reports have mentioned library involvement in managing research data, but often only in general terms or grouping librarians with other professionals without differentiating their respective roles. For example, the US National Science Board report on *Long-Lived Digital Data Collections* categorises librarians along with several other specialists (information and computer scientists, database and software engineers and programmers, disciplinary experts, curators and expert annotators, archivists) as 'data scientists' (NSB, 2005, p. 27). The NSF (2007, p. 38) also suggested that professional education for careers in data management and curation may need 'new, hybrid degree programs that marry the study of library science with a scientific discipline'.

Within the library world, the US Association of Research Libraries (ARL) has actively promoted the role of libraries in data management and e-science through a succession of reports on the subject that envisage a significantly expanded role in digital data stewardship for research libraries in collaboration with other stakeholders. Key themes highlighted by the ARL include the importance of partnerships, the challenge of working across disciplinary and institutional cultures and the need for education and outreach (Friedlander and Adler, 2006; Lougee et al., 2007). A survey in 2009 revealed overall enthusiasm among ARL members for new roles in the academic research process and identified 21 libraries (from 57 replies) already providing infrastructure or support services for e-science, with another 23 planning to do so. The findings confirmed the importance of collaboration, but also highlighted the relevance of library and information science (LIS) expertise (Soehner et al., 2010).

Within the UK, data-related library activity is largely represented by the local data libraries/support services established at the Universities of Edinburgh and Oxford in the 1980s and the London School of Economics in the 1990s, whose focus has been on social science data and geographical information systems (Macdonald and Martinez, 2005). There have been some studies exploring national and local approaches to the management of research data, including a feasibility study for a national research data management service instigated by Research Libraries UK and investigations of the use of repository technologies for data curation (Lewis, 2010; MacDonald and Uribe, 2008; Martinez-Uribe, 2007). A few libraries have been involved in data management projects funded by JISC, such as the Institutional Data Management Blueprint at the University of Southampton (Takeda, 2010). There has also been significant debate on the roles of librarians and other stakeholders in the management of research data and on their skills, training and development needs (Lewis, 2010; Lyon, 2007; Pryor and Donnelly, 2009; Swan and Brown, 2008).

A more recent ARL report offers a bolder vision of new roles in digital curation, prefaced by the assertion that 'the strongest future for research libraries is one in which multi-institutional collaborations achieve evolvable cyberinfrastructures and services for digital curation' (Walters and Skinner, 2011, p. 5). This report anticipates a shift away from the traditional

public (front-of-house) and technical (back-of-house) service mindset towards 'the trio of strong infrastructures, content, and services', where infrastructure is used to mean not only facilities and technologies, but also human expertise (Walters and Skinner, 2011, pp. 5-6). Box 1 shows the proposed 'new roles' for research librarians in digital curation. Some of these roles look novel, but others are more familiar, with the novelty derived from the *context*, rather than the focus of the role.

- Acquisitions and Rights Advisers
- Teachers/Instructional Partners in Learning Spaces
- Observers/Anthropologists of Information Users and Producers
- Systems Builders
- Content Producers and Disseminators
- Organizational Designers
- Collaborative Network Creators and Participants

Box 1 – New Roles for Digital Curation (Walters and Skinner, 2011)

**Data management and library practice**

The debate about library involvement in data management has been conducted at both strategic and operational levels. Lewis (2010, p. 145) has previously argued that 'data from academic research projects represents an integral part of the global research knowledge base, and so managing it should be a natural extension of the university library's current role in providing access to the published part of that knowledge base', while also noting 'the scale of the challenge in terms of infrastructure, skills and culture'. Engaging with a a completely new area of practice is clearly a challenging prospect, irrespective of the strategic case for getting involved, especially at a time when library budgets worldwide are under acute and continuing pressure as a result of the global economic downturn (Harper and Corrall, 2011; Nicholas et al., 2010). However, even though research data sets are in several ways quite different to other library resources, practitioners working in the area have identified strong links and significant overlaps with existing library interests and activities.

First, the open data movement associated with e-science and e-research in general is the natural culmination of the open access movement, open source software and open standards, which all support the traditional commitment of libraries to the free flow of information and ideas (Corrado, 2005; Lougee et al., 2007; Macdonald and Uribe, 2008; Willinsky, 2005). Librarians are already active proponents of open access, typically taking the lead in establishing and developing institutional repositories within their universities, generally in collaboration with computing/IT services and other units (Rieh et al., 2007). The International Federation of Library Associations and Institutions (IFLA, 2011) has recently endorsed open access as an 'essential issue' within its information agenda, highlighting the work already done by librarians in building infrastructure, creating user-friendly services, educating research stakeholders, helping with deposit of outputs and securing long-term access.

Significantly, the IFLA statement explicitly includes 'research data curation and sharing' among the open access support provided by librarians (IFLA, 2011, p. 2). Walters (2009, p. 84) argues that data curation programmes represent a 'robust growth trajectory' for library institutional repository initiatives, which was the case for Georgia Tech Library and

Information Center. Interestingly, Walters (2009) emphasises that the programme initiated at Georgia Tech, and similar 'entrepreneurial steps' taken by like-minded colleagues at six peer organisations in the US (Johns Hopkins University, University of California-San Diego, University of Illinois at Urbana-Champaign, University of Michigan, Cornell University and Massachusetts Institute of Technology) have essentially been bottom-up initiatives where individual library and technology professionals have reached out to faculty and research centres 'without the benefit of national mandates and high-level university policies'.

Secondly, the characterisation of e-science as 'a new mode of "collection-based" research' (Berman et al., 2002, p. 37) or 'collection-based science' (Beagrie, 2006, p. 5) also suggests a strong connection with library practice. Although institutional repositories are often established as distinct digital library or information management initiatives, Genoni (2004) argues that the management of repository content is essentially a *collection* management issue. Witt (2008) highlights the relevance for data curation of library expertise in classification and description through cataloguing and metadata, as well as experience in selecting, deselecting and presenting information in an appropriate context. He also notes that many research libraries have special collections staff with expertise in the appraisal and preservation of primary source materials. Choudhury (2008, p, 218) similarly sees parallels between new roles in managing research data and established responsibilities for preserving and curating special collections, suggesting that 'scientific datasets may be thought of as the "special collections" of the digital age'.

Similarly, the ARL E-Science Task Force notes that the dual focus of data curation on both preservation and active management for access fits well with the mission and concerns of research libraries (Lougee et al., 2007). The ARL Task Force has also argued that library expertise in developing systems and standards for digital content and associated services can be exploited to create discovery and management systems for digital data, pointing to library experience with both institutional and disciplinary repositories, with integration and interoperability tools, and with business and technical aspects of long-term archiving, though the report also points out that existing expertise and infrastructure 'will be seriously stretched by the new, more complex demands of e-science' (Lougee et al., 2007, p. 6). Salo (2010) usefully draws out both similarities and differences in the infrastructure (technological and human) needed to manage research data, warning against naïve repurposing of systems and staff, but concluding positively:

> 'None of the challenges presented herein should discourage librarians from engaging with the research data challenge. Our unique expertise in metadata, digital preservation, public service, and technology translation will serve researchers well, as will our sturdy common sense and the domain expertise of our subject librarians.'

Education and training for both specialist and non-specialist users of datasets is a recurring theme of official reports on digital data collections. The NSB (2005, p. 27) assigned responsibility to data scientists (including information scientists, librarians and archivists) for 'education and outreach programs that make the benefits of data collections and digital information science available to the broadest possible range of researchers, educators, students, and the general public'. Carlson et al. (2011, p. 630) suggest that the widespread involvement of academic libraries in information literacy education makes the development of an educational role in data management and curation 'a logical entry point into increasing libraries' role in supporting e-research' and Gabridge (2009, p. 17) similarly suggests that subject liaison librarians can easily extend their instruction roles 'to help student researchers understand what to do with their data and increase their awareness of library resources'.

Data-related reference and consultation services are also seen as 'a natural fit for subject librarians who provide similar services for other types of information' (Soehner et al., 2010, p. 16). Examples mentioned include identifying datasets to meet student or faculty needs, providing access to data resources and advising researchers on current standards for organization of data in specific subject areas, in addition to help with the specific task of developing data management plans and more general awareness raising through creation of special websites to describe services available. Recent policy decisions by research funding bodies to require researchers to submit plans for data management and sharing with their grant applications have given librarians a timely opportunity to offer information on funders' requirements and assistance with constructing a plan. The ARL has provided a set of web pages to help librarians make sense of the new NSF requirement (Hswe and Holt, 2010). The UK Digital Curation Centre has also provided a comprehensive set of resources for data management planning, which includes a summary of UK funders' data plan requirements, in addition to a checklist, guidance, examples and a web-based planning tool (DCC, 2010).

**Strategies for engaging with data**

Our review has shown how the management of research data resonates with library values (such as open access) and can be related to current operations and professional practices in several areas, including institutional repositories, collection management, systems development, information literacy and reference services. We have also pointed out a significant opportunity for librarians to offer valuable support for research in their institutions, by assisting investigators with data management planning and thus help them fulfil new requirements of funders to submit plans for data sharing with their grant proposals. Libraries that have not yet engaged with the research data agenda need to review possible areas of activity and decide where investment of effort is likely to have the most benefit for their particular institution, taking into account the local and national context. Working in partnership with other campus agencies to determine where to begin is recommended, notably computing/technology services, research offices and those responsible for research governance (such as a Pro-Vice Chancellor or Vice Provost for Research).

Lewis (2010, p. 154) previously suggested nine areas of strategic, tactical and operational engagement in research data management, grouped in a pyramid for ease of reference. We can now add data collection development as another area of activity discussed in recent library literature (Newton et al., 2010). Figure 1 shows the research data management pyramid for libraries supported by collection development and data resource management as a base layer of activity, represented here by the key processes of identifying, selecting, describing, preserving and presenting data resources for use. This list does not cover the complete set of 'full lifecycle', 'sequential' and 'occasional' actions specified by the DCC Curation Lifecycle Model (Higgins, 2008), which is often used to define the work involved in managing research data (Harvey, 2010; Lewis, 2010), but concentrates instead on areas where librarians already have skills and expertise that can be applied to research data curation (Witt, 2008).

Libraries engaging with research data for the first time can learn from the experiences and services of pioneers in the field, discussed in the growing body of academic and professional literature on the subject and also described on their institutional web pages. The websites of the ARL and DCC offer additional guidance. We highlight some prominent examples here.

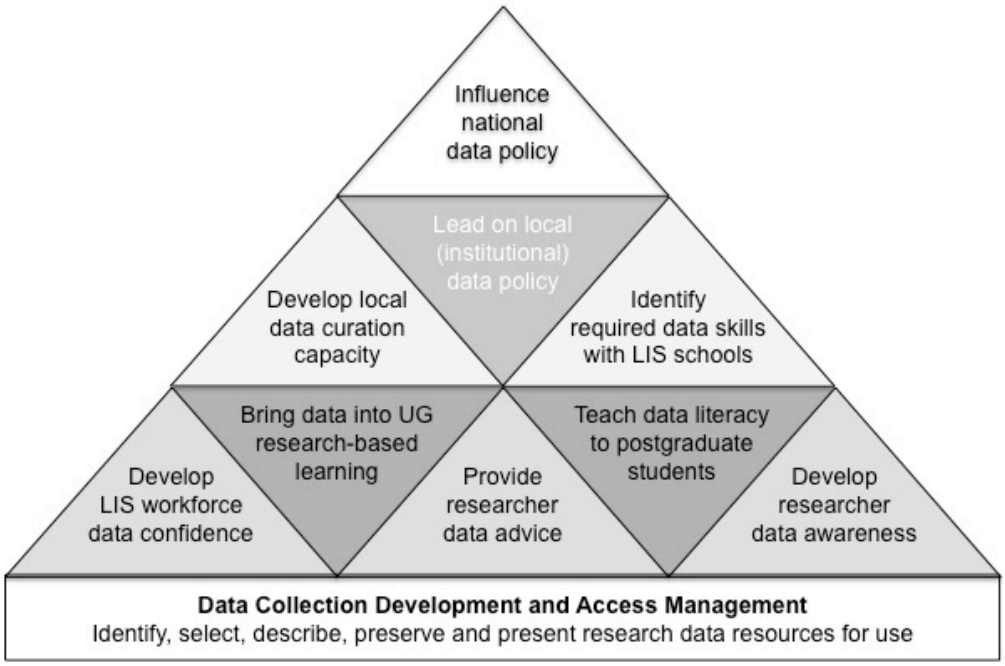


Figure 1 – Research Data Management Pyramid for Libraries

***Influence national data policy***

Librarians in the US have worked through the ARL to influence national developments in research data management. For example, the NSF not only funded the workshop organised by the ARL on Long-term Stewardship of Digital Data Sets (Friedlander and Adler, 2006), but has also acted on key recommendations, by funding data curation projects involving libraries working in partnership with other stakeholders and by requiring data management plans to be included in grant proposals. In Canada, the research library community is strongly represented on the Government's Research Data Strategy Working Group (Canada, 2011) and similarly in Australia, senior members of the library community have been appointed to the Steering Committee of the Australian National Data Service (ANDS, 2011). The European Union High Level Expert Group on Scientific Data also included the Head Librarian of CERN, the European Organization for Nuclear Research, among its members (HLG, 2010).

***Lead on local (institutional) data policy***

Library directors and other senior library staff have a key role to play in helping senior administrators to understand the nature and importance of the data management challenge and develop an appropriate and coherent response. This will typically mean working with institutional research committees and other senior officers to achieve a joined-up approach. It may involve formulating a data management policy and/or making a business case for investment in research data management. For example, the University of Edinburgh has developed a policy statement on research data management, which is supported by a background document that includes links to a checklist and advice on data management planning available from the Information Services website (University of Edinburgh, 2011). Edinburgh's Policy on Management of Research Data was developed by a Research Data Management Working Group convened by the Director of Library and Collections.

***Develop local data curation capacity***

Several libraries in the US have used involvement in institutional repository management as the basis for exploring data curation via dataset deposit, notably Purdue University (Witt, 2008) and Georgia Tech (Walters, 2009), although extending repository services from publications to datasets is easier to conceptualise than to achieve technically (Salo, 2010). Financing scalable storage is another challenge: both Georgia Tech and Purdue are looking at cost-recovery models in collaboration with their campus technology services. Three UK universities have also explored dataset repository development via the DISC-UK DataShare project, experimenting with different software platforms (DSpace, ePrints and Fedora) and different architectures, with two incorporating datasets in their publications repository, the third developing a separate interoperable repository (Rice, 2009). The DataShare team has produced a comprehensive guide based on experience and research to inform local decision-making and planning for expansion of repository services (Green et al., 2009).

***Identify required data skills with LIS schools***

National and international reports on research data management have flagged the need for development of education and training programmes for the professionals expected to take on responsibility for managing research data in their institutions (NSF, 2007; HLG, 2010). Swan and Brown (2008, p. 30) recommended that the UK research library community work with other stakeholders 'to develop a curriculum that ensures a suitable supply of librarians skilled in data handling', sensibly recognising that educators need a steer from employers here on both curriculum content and market demand. In the US, a recommendation from ARL that the NSF and the Institute of Museum and Library Studies (IMLS) should fund training to prepare current and future library and information professionals for credible roles in data stewardship (Friedlander and Adler, 2006) has been followed through in a succession of innovative development projects funded by IMLS (Hand and Davidson, 2009; Ray, 2009). We discuss education and training more fully later in the chapter.

***Develop LIS workforce data confidence***

A survey of the UK academic library sector in autumn 2006 commissioned by a CURL/SCONUL Task Force on e-Research found poor awareness of activities in the field among academic liaison/subject librarians, apart from institutional repository developments (Martinez, 2007). Research data have arguably gained a higher profile recently in the library community as a result of reports and events sponsored by the JISC and DCC, in addition to publications on the subject aimed at library practitioners (Lewis, 2010; Macdonald and Uribe, 2008; Salo, 2010; Westra, 2010), recommended readings on e-science (Szigeti and Wheeler, 2011) and a useful webliography on data curation, covering email lists, journals, guidelines, reports, organisations, directories, standards and software (Westra et al., 2010). The New England eScience Portal for Librarians is another useful educational resource offering overviews and primers (University of Massachusetts, 2011). However, managers may need to point library staff towards such sources to help them update their knowledge and gain the confidence needed to hold conversations on the subject with researchers.

***Bring data into undergraduate research-based learning***

American and European reports have called for data handling to be embedded in university and school curricula to support public understanding and participation in science (HLG, 2010; NSB, 2005). Carlson et al. (2011) argue that librarians should develop a data information literacy curriculum, working with disciplinary faculty to define the skill sets needed and to teach the skills identified. Ogburn (2010, p. 244) urges librarians to provide programmes for both scientists and non-scientists 'that teach the interpretation of data and

visual representations of research findings'. Many academic libraries provide specialist data resources and services to support research and teaching in social science disciplines (Czarnocki & Khouri, 2004; Read, 2007) and some have developed courses for undergraduates in the use of statistical data resources (Stephenson and Caravello, 2007). Extending provision to enable undergraduates to work with real research data fits well with the problem-based/inquiry-based learning pedagogies favoured by research universities.

MacMillan (2010) provides a detailed case study showing how genetic data resources were incorporated into a third-year information literacy lab at Calgary University that simulated the working methods of experienced researchers. A structured step-by-step exercise based on tool-specific modules provided demonstration, practice and discussion of a set of resources, progressing from online encyclopedias and journal databases through *Google Patents* to gene and protein databanks and tools, which enabled the liaison librarian to highlight synergies and relationships among key resources. Macmillan (2010) argues that authentic assignments designed in collaboration with discipline faculty are the best way to foster student familiarity with key tools in the field and that 'faculty appreciate library partners who understand the structures of information in the discipline that complement more traditional bibliographic tools'. He provides full details of the questions set for each task and invites other librarians to use his resources to develop their undergraduate teaching.

Carlson et al. (2011, p. 633) use the term 'data information literacy' to distinguish their conception of the competencies needed for e-research from what they see as narrower prior concepts of literacy, explaining that their concept 'merges the concepts of researcher-as-producer and researcher-as-consumer of data products' and 'builds upon and reintegrates data, statistical, information, and science data literacy into an emerging skill set'. Based on interviews with research faculty at two universities, assessments of advanced undergraduate and graduate students on a Geoinformatics course at Purdue and comparison with the ACRL (2000) *Information Literacy Competency Standards for Higher Education*, they propose a set of thematically arranged core competencies as generic educational objectives for a data information literacy programme. They note that the balance between curriculum components and also between specific content elements will vary for different disciplines. Each theme consists of a heading supported by a description in two or more sentences (Carlson et al., 2011, pp. 652-653). Box 2 lists the 12 thematic headings.

- Databases and Data Formats
- Discovery and Acquisition of Data
- Data Management and Organization
- Data Conversion and Interoperability
- Quality Assurance
- Metadata
- Data Curation and Re-use
- Cultures of Practice
- Data Preservation
- Data Analysis
- Data Visualization
- Ethics, including citation of data

Box 2 – Core Competencies for Data Information Literacy (Carlson et al., 2011)

***Teach data literacy to postgraduate students***

Most UK university libraries have either formal or informal involvement in research skills training for postgraduate students, though few currently cover research data management (RIN, 2008). Postgraduate research training is a priority area for library efforts in supporting the management of research data as it presents an important opportunity to influence the behaviour of future researchers by developing their understanding of the generation, description, preservation, storage and use of data in line with good ethical and technical practice. UK information literacy practitioners have recognised the need to develop their teaching in this direction and have engaged with the postgraduate training agenda at a national level: first, by influencing the information and data literacy content of the Researcher Development Framework, the new national competency tool for researchers (Vitae, 2010); and secondly, by developing a *Research Lens* that specifies data-related understanding and abilities for the revised and expanded version of the *Seven Pillars of Information Literacy* (SCONUL, 2011).

The Seven Pillars model is the most widely used information literacy framework in the UK and has been adopted by librarians and teachers around the world to support information skills training and information literacy development. The new format enables more detailed specification of information literacy attributes with explicit tailoring to the needs of different groups of information users. The introductory section of the *Research Lens* explains that information literacy 'encompasses concepts such as…data curation and data management' (SCONUL, 2011, p. 3). Box 3 gives examples of data-related abilities included under each of the *Seven Pillars.*

*Identify* – recognise a need for information and data to achieve a specific end and define limits to the information need

*Scope* – identify which types of information (e.g. data, people, videos, published information) will best meet the need

*Plan* – select the most appropriate search tools (people, search engines, databases etc.) and data collection techniques

*Gather* – access full text information, both print and digital, read and download online material and data

*Evaluate* – assess the credibility of the data gathered

*Manage* – Identify data curation opportunities to ensure that research data is ethically stored for re-use in other projects

*Present* – see connections between sections of own data and the literature

Box 3 – Data-handling Abilities of information Literate Researchers (*Extract from* SCONUL, 2011)

***Provide researcher data advice***

Many US librarians have extended their reference and consultation services to offer information and advice on issues related to data management, often building on existing offerings related to open access and other aspects of scholarly communication, including copyright and intellectual property, metadata and technical standards, data archiving and preservation. Several libraries provide curation services jointly with supercomputer or research computing centres, with the computer centre providing computation, storage and

back-up services (for example, Cornell University and University of California San Diego). Research data services for digital curation at the University of Wisconsin-Madison are provided by 'a group of librarians, IT staff, and graduate students in the School of Library & Information Studies' (UW-Madison, 2011a). Box 4 shows the description of service offerings presented on the UW-Madison website.

## Our Services

All our services are free to UW-Madison faculty, researchers, staff, and graduate students.

### Data management plan help

• We can help **draft a plan** to meet requirements from NSF and other funders.
• We can also **review your plan** and suggest improvements.

### Consultations

- Data **workflow and process improvement** in your department, research unit, or laboratory.
- **File-format and metadata standards** that fit your research and your community.
- **Digital preservation and archival** concepts, to help you avoid losing your work.
- Advice on **data sharing and reuse rights**, to maximize your influence and credit.
- **Database design advice** and **data modeling suggestions** to get the most from your data.

### Training and education

- We will **train your trainers** in data-management best practices.
- We also **train you and your lab**, customizing our approach to what *you* want to accomplish.
- We come to **research-methods courses** to train the next generation of researchers.
- We bring our expertise to your **symposium, brown-bag, or meeting**.

### Referrals

- **Storage and backup solutions**, on campus and off.
- **Data-security experts**, particularly in the Office of Campus Information Security.

See something you need? Fill out the form below and we'll be in touch.

Box 4 – Research Data Services at UW-Madison (UW-Madison, 2011b)

Hswe and Holt (2011) report a huge surge of library activity around data management planning in the wake of the announced NSF requirement, in the form of web pages, templates, tutorials, webinars and workshops, though some libraries were already providing services in this area before the announcement. In a few cases, library data management planning web pages consist entirely of links to resources on other institution's websites, showing how easy it is now for libraries to develop a web presence in this area. Notable examples of US data management planning resources include:

- MIT Libraries' web page on *Data Management Plans* (part of a set of a pages on Data Management and Publishing), which lists recommended components for plans and provides links to external web-based resources, as well as offering an email address for help (MIT, 2011)
- University of Connecticut Libraries' web page on *Data Management Plans for Grant Funded Projects* (part of a Scholarly Communications website), which also lists typical elements of plans, but at a more detailed level, again providing links to additional web resources (University of Connecticut, 2011)
- Purdue Libraries' *Data Management Plan Self-Assessment Questionnaire*, which was developed by adapting data collection/curation tools already developed and contains 31 questions under seven headings: Describing the research data, Data Standards, Metadata Standards, Data Sharing, Data Access, Intellectual Property and Re-Use, Data Archiving and Preservation (Purdue University Libraries, 2011).
- Duke University Libraries' LibGuide on *Research Data Planning,* which is presented in the same format as their subject-based resource guides, enabling users either to navigate a comprehensive set of web pages to find specific information or to print the complete guide as a single document (Denniston *et al.,* 2011).

Many other libraries have used the standard LibGuide template to provide guidance both on data management planning and other data-related matters, for example on data curation and archiving or on finding data resources; other guides in this format provided by Duke University Libraries include *Scientific Data Repositories and Datasets* (Gray, 2011).

### *Develop researcher data awareness*

In addition to providing specific information and advice, libraries can play an important role in general awareness-raising and advocacy around data management for their research communities, whose level of understanding and interest is likely to vary across campus. Shearer and Argáez (2010) advise that even general messages should ideally be tailored to particular groups, because of the different disciplinary cultures that often make the issues and challenges of data stewardship discipline-specific. Librarians with limited knowledge and experience of data management can draw on the array of web-based resources now available from centres of expertise to support this work, such as the Digital Curation Centre's series of introductory briefing papers and the Australian National Data Service's set of awareness guides, as well as reviewing material on other libraries' websites.

### *Develop and manage access to data collections*

Traditional conceptions of a library collection have been expanded over the past three decades to incorporate both tangible physical and intangible digital objects that may be stored locally or remotely and owned by the library as a result of purchase, borrowed or managed by the library on behalf of others, accessed by library members under licence or selected for promotion to users from public domain resources (ARL, 2002; Gorman, 2000). Digital repositories and data resources (deposited in remote disciplinary or local institutional repositories) are recognised as part of this framework (Lynch, 2003), which has been variously depicted in concentric circles (Lee, 2003, p. 432) or a 'collections grid' (Dempsey, 2003, p. 124). Modern collection development is less about acquiring and hosting materials and more about negotiating and facilitating access to resources, but selecting and describing

the information sources offered to users are still key activities for the library in the 21st century (Brophy, 2007).

Collection development and access management for data resources thus potentially includes building local data collections (which may be hosted by the library or stored in other institutional locations) and providing access to remote publicly-available free, licensed or subscription-based data products and services. Many libraries already have experience of developing collections of externally-sourced data that can inform policy and practice for data collection development based on locally-produced datasets, for example through subscribing to statistical series (in hard-copy and electronic services) and developing geospatial data collections based on government, commercial or academic data (Florance, 2006; Humphrey and Hamilton, 2004). The tasks involved in building locally and externally sourced collections are similar, such as identifying potential material for selection and describing the chosen resources for presentation to users; but the procedures are different, requiring application of knowledge and skills in new ways and/or development of new skills (Newton et al., 2010).

Several authors have suggested that libraries can play particularly valuable role in collecting and archiving smaller-scale datasets generated by research groups or individual academics whose disciplines lack national or international repository infrastructure (Hey and Hey, 2006; Luce, 2008). Commentators also stress the need for librarians to reach out to academics proactively and become involved as partners 'in the early planning and data-modeling phases of eResearch' (Luce, 2008, p. 46) to ensure lifecycle management of data so that they can be preserved and used. Borgman (2010, p. 4) also argues that 'early engagement is key to taking a temporal approach' and the need to reconfigure library organizations 'from being a reading room to a full research center'. As Newton et al. (2010) point out, methods such as publisher catalogues/feeds cannot be used to identify potential data acquisitions, so data selectors need to turn instead to faculty research profiles and grant announcements from academic departments and research offices.

The DCC and ANDS have provided a 'working level' guide to the appraisal and selection of research data for curation, which explicitly identifies a major role for information professionals in formulating a policy for their institutions (Whyte and Wilson, 2010). The guide identifies seven general criteria for assessing the value of a dataset, which are then elaborated as a checklist of bullet points under a series of questions. Newton et al. (2010, p. 67) provide a real-world example in a table of 'collection criteria' with supporting rationale developed at Purdue University Libraries, covering similar points, but adding an additional requirement for the data to have an 'Institutional association', defined as 'a clear institutional connection to Purdue University'. This point is more specific than Whyte and Wilson's (2010) 'Relevance to mission', a requirement that is covered separately by Purdue under the heading 'Value to Purdue's collection', which refers explicitly to the institution's 'research, teaching, and discovery missions'. Table 1 compares the two sets of criteria.

Table 1 – DCC/ANDS and Purdue Selection Criteria for Data Collections

| **DCC & ANDS** (Whyte and Wilson, 2010) | **Purdue University Libraries** (Newton et al., 2010) |
|---|---|
| | Institutional association |
| 1. Relevance to mission | Value to Purdue's collection |
| 2. Scientific or historical value | Value to research or education generally |
| 3. Uniqueness | Uniqueness and availability of the data |
| 4. Potential for redistribution | Degree of restrictions placed on the data |
| 5. Non-replicability | Format of the data |
| 6. Economic case | Cost |
| 7. Full documentation | Condition of the data and its documentation |

At a practical level, several libraries report the use of 'data interviews' to identify datasets for curation (Walters, 2009; Witt, 2008), which can be seen as an upstream repurposing of the traditional library *reference* interview in terms of the process and skills used. Practitioners at Purdue University Libraries have devised a set of questions (displayed on a poster) that can be used by subject/liaison librarians to interview researchers and explore the suitability of a dataset for adding to the library/institutional collection. Witt and Carlson (2007) argue that the process of evaluating research datasets for selection as information assets and exploring the infrastructure and services needed for their preservation and access is similar to traditional collection development practices, although there are obviously differences in the formats of the material being considered and also its status that make the process more complicated, hence their suggestion that an interview is an effective method for identifying and selecting potential additions to stock. Box 5 shows their set of questions, which are explained in more detail on their poster.

1. What is the story of the data?
2. What form and format are the data in?
3. What is the expected lifespan of the dataset?
4. How could the data be used, reused, and repurposed?
5. How large is the dataset, and what is its rate of growth?
6. Who are the potential audiences for the data?
7. Who owns the data?
8. Does the dataset include any sensitive information?
9. What publications or discoveries have resulted from the data?
10. How should the data be made accessible?

Box 5 – Interview Questions for Data Selection in Collection Development (Witt and Carlson, 2007)

These questions may lead on to more specific supplementary questions, notably on the potential uses and audiences for the data, which are identified as primary selection criteria for datasets to be added to a library collection (Witt and Carlson, 2007). The Purdue team has also provided comprehensive step-by-step guidance on conducting data interviews in an 'Interviewer's Manual' and an 'Interview Worksheet' (for completion by the interviewee) as part of their *Data Curation Profiles Toolkit,* which can be downloaded from the Data Curation

Profiles Community website (Purdue University Libraries, 2010). The tools were developed as a result of a project investigating 'Which researchers are willing to share data, when, with whom, and under what conditions?' carried out by the Libraries with the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign with funding from IMLS, which has also funded promotion and development of the toolkit through a national programme of training and evaluation for library and information professionals.

**Professional education for data management**

As research data become an increasingly important part of the information landscape and an essential dimension of the knowledge base supporting not only future research, but also learning, teaching and evidence-based policy and practice, library and information professionals in all sectors must be prepared to facilitate access to various types of datasets, irrespective of the role played by their own library or other information organisation in managing research data. Providers of professional education for library and information specialists have a key role to play here as part of their ongoing responsibility to ensure that the content of their programmes, courses and modules is continually reviewed and updated to anticipate change and reflect developments in the service environment. Education and training for data-related library activities needs to be provided for both new professionals and experienced practitioners to enable libraries to fulfil their potential and also to ensure that institutions do not assign responsibilities to others who have relevant subject expertise and/or technology know-how, but lack the informational, managerial and personal abilities that are essential to apply the desired specialist competencies successfully.

Many postgraduate LIS programmes include units on digital libraries and/or digital technologies as distinct elements of their curricula (Ray, 2009), but others have simply embedded digital developments into existing courses. Existing provision has been criticised for its rudimentary treatment of both theoretical and practical aspects (Varalakshmi, 2009) and insufficient opportunities for hands-on practice (Dahlstrom and Doracic, 2009). Current coverage of data management and digital curation in postgraduate curricula is limited and uneven across the sector. Within the UK, Loughborough University is the only information science department offering an elective course in Digital Curation as part of a postgraduate library programme, though there is more coverage of digital curation/preservation in other types of programmes, such as those in records management at the University of Dundee, information management at the University of Glasgow and the digital asset management at King's College, London (Pryor and Donnelly, 2009).

Provision in the US is more extensive, where many grants from the IMLS over the past five years have enabled LIS schools to develop and launch new specialised postgraduate courses and concentrations (pathways) that can be taken as part of or an add-on to standard Master's programmes (Hank and Davidson, 2009; Ray, 2009). Notable examples of initiatives funded by IMLS cover both professional preparation and continuing education programmes:

- The University of North Carolina at Chapel Hill School of Information and Library Science DigCCurr project has developed an openly-accessible graduate level curriculum to prepare master's students for work in digital curation and an international doctoral curriculum to prepare future faculty to teach and research in the field (Hank and Davidson, 2009). The objective was to 'build modules around topics, rather than construct entire courses, so that faculty can develop their own classes using the modules, and students can more easily develop courses of study around their own interests' (Ray,

2009, p. 364). The project also included the development and delivery of annual, multi-stage training institutes (five days followed by two days six months later) for professionals already working in the area, aiming to 'build an international community of digital curation practitioners', as well as developing their knowledge and skills. The 2010 Institute attracted 33 participants from academic libraries and archives, non-profit consortiums and corporations, government archives and public libraries (Costello and Brown, 2010).

- The University of Illinois at Urbana-Champaign Graduate School of Library and Information Science has developed a Data Curation Education Program (DCEP) as a four-course concentration (a designated pathway) within its existing MS in LIS. In addition to required courses on data curation, digital preservation and systems analysis, DCEP includes a period of field experience/internship, emphasising the role of experiential learning (Ray, 2009). An initial focus on science data curation was later expanded to cover the humanities and the project also included delivery of a week-long summer in-service training institute in humanities data curation for 24 established professionals, which attracted an interesting mix of academic librarians that included subject or area specialists, cataloguing or metadata librarians, digital library/content specialists and data librarians. The Institute was over-subscribed and generated requests for additional in-service education adapted to specific roles and at different levels (Renear et al., 2010).
- The University of Michigan School of Information has established ten 360-hour summer internships at organisations with strong digital curation programmes for students taking a new Preservation of Information (PI) specialisation within its MS in Information. The PI specialisation offers ten courses, in addition to the internship and a doctoral-level seminar on data curation (available to advanced master's students). The internship design was 'based on a cognitive apprenticeship model with the goal of strengthening each student's ability to think like a digital curator'. The funding also enabled the creation of a special internship course, Practical Engagement Workshop in Digital Preservation, to provide instruction related to the students' projects, including organisational skills. Although the PI specialisation originated as an alternative concentration to Archives and Records Management (ARM), students may take it in addition to either the ARM or Library and Information Services specialisation, which illustrates the growing convergence of interests among different professional groups (Yakel et al., 2011, p. 30).
- The University of Arizona School of Information Resources and Library Science has developed an online six-course graduate certificate in Digital Information Management in partnership with the Arizona State Library, Archives and Public Records. DigIn aims to provide an interdisciplinary learning experience to prepare librarians, archivists, records managers, software developers, systems administrators and science data managers for 'boundary-spanning, technology-intensive roles as collection managers in their respective professional communities'. Although it has a 'strong hands-on element' that is intended to promote technical skills, its declared focus is not on specific technologies *per se,* but rather on 'technological and systems thinking' for 'technological fluency', with an emphasis on 'the learning skills needed to evaluate and work effectively with new technologies as they emerge'. The final course is a semester-long period of field experience working on a capstone project in an organisation supervised by a practitioner (Fulton et al., 2011, pp. 96-98).
- Syracuse University School of Information Studies has developed an eScience Librarianship specialization within its MS in LIS in partnership with Cornell University Library to train students with a science or technical background as a new generation of 'eScience Librarians'. The digital curation curriculum for the eSLib project builds on work

done in other projects, but emphasizes the management and preservation of science-related information. The required courses include Scientific Data Management, Cyberinfrastructure and Scientific Collaboration, Database Systems, Metadata, and Data Services (capstone). The eSLib Fellows are paired with Library mentors and participate in special events and regular working group sessions at the Library. They are also involved in smallscale research projects with Syracuse researchers and librarians, and In addition benefit from paid work experience, a paid summer internship and funded travel to various eScience-related conferences and events (Qin et al., 2010; Syracuse University, 2010).

### *Key issues in data management education*

Despite the noticeable diversity in provision, some common themes emerge from published accounts of curriculum development for data management and curation. First, the importance of emphasising ***digital lifecycle*** stages in the core curriculum is often mentioned as one of the key factors differentiating management and curation of digital data from stewardship of legacy library materials (Gregory and Guss, 2010). This is evident in the way that the DCC Curation Lifecycle Model (Higgins, 2008) has been used as a framework for instruction in courses and handbooks. For example, the DigIn programme team 'decided to offer one course that essentially covers the "curate" half of the Digital Curation Centre's Curation Lifecycle Model,and a second course dedicated to the "preserve" half of the model' (Botticelli et al., 2010, p. 4). Similarly, all the chapters in Parts 2 and 3 of Harvey's (2010) *Digital Curation: A How-To-Do-It Manual* are structured respectively around the Full Lifecycle and Sequential Actions of the DCC Model.

Second, the requirements for and role of ***technology*** in the digital curation curriculum are frequently mentioned. Commentators invariably highlight the need for a solid technological infrastructure, discussing problems and solutions for both students and instructors (Fulton et al., 2011; Yakel et al., 2011), but they also stress that fundamentally their courses are not about developing advanced *technical* skills; their goals focus instead on *learning* skills. Thus at Arizona, 'our goal was to enable a broad range of information professionals to gain an essential level of fluency or literacy with the advanced technologies needed for curation, especially server-based repository applications and Web-based content management systems' (Fulton et al., 2011, p. 97) and at Michigan, 'Our goal is not to teach the tool, but to use the tools to enable students to learn how to learn new tools' (Yakel et al., 2011, p. 28).

Third, the value of practical field experience, typically ***internships*** with real projects to work on, is arguably the strongest and most consistent message from the cases examined (Hank and Davidson, 2009; Qin et al., 2010; Ray, 2009; Yakel et al., 2011). Practical placements are an established feature of postgraduate LIS programmes in the US, though less common in the one-year UK master's degree, but clearly have particular value when students are being introduced to an emerging area of practice that lacks an established body of knowledge and supporting literature. When Stanton et al. (2011, p. 90) evaluated the experiences of five Syracuse library/information master's students placed in science research centres by analysing detailed logs of their activities and exit questionnaires rating changes in their capabilities, they concluded that a targeted internship of this type should be a 'critical part' of any programme for eScience professionals, identifying several key competencies that would be much harder to develop effectively in the classroom (e.g. 'Analyzing project or researcher needs'). Direct evidence from two graduates of the North Carolina DCCurr project reinforces this view:

> 'Both fellows agreed that their practicum experiences and master's paper research projects were pivotal in reinforcing and enhancing topics covered in

> coursework. Whereas particular courses often touched on select functions, the master's papers and practica aggregated an entire spectrum of concepts and brought theoretical discussions into real life institutional contexts.' (Gregory and Guss, 2010, p. 9).

The digital curation courses and programmes developed to date are aimed at students and practitioners intent on roles with a strong focus on digital curation. However, if research datasets really are to become mainstream resources for library and information professionals, they need to be covered explicitly in the core LIS curriculum, as well as being the subject of specialist courses, so that all new professionals gain at least a basic understanding of their value and use, which will also help them to make informed decisions on whether to take any relevant specialist courses offered. Good practice demands continual review and regular updating and development of all courses, lectures, activities, readings and assignments to ensure that teaching incorporates the latest thinking in the discipline and practice in the field. In the context of managing research data, educators must ensure that there is appropriate coverage of this emerging area of practice in courses covering topics such as collection development and management, reference and information services, information literacy and instruction, academic liaison and support, and library systems and technologies. If courses involve practitioners as guest lecturers, course leaders also need to check that they are prepared to discuss their involvement with research data.

Incremental curricular developments of this kind are rarely reported in the literature and are hard to track as they often do not require changes to course descriptions, but most professional educators are continually refreshing their modules. For example, at the University of Sheffield, our core unit on Information Resources & Information Literacy now includes primary research data among the resources discussed in the Week 2 lecture session on 'The information universe' as an example of emergent types of reference material. Similarly, the introductory lecture on 'The concept of the collection and collection management' in our core unit on Libraries, Information & Society emphasises that a library 'collection' may now encompass born-digital resources such as geo-spatial and numeric datasets along with the tools for their management and use (ARL, 2002), using the OCLC Collections Grid to show how research data created within an institution needs to be managed alongside externally-sourced content (Dempsey, 2003). Our Academic & Research Libraries elective extends this coverage to other types of data and introduces students to the DCC Curation Lifecycle Model (Higgins, 2008) in the context of e-research, involving a practitioner with experience in repository development for this session.

Specialist courses on new subjects also need continual review, particularly in emerging fields where practice is not well established. Borgman (2010) reports that as soon as a new course on Data, Data Practices, and Data Curation at the University of California, Los Angeles (UCLA) had been delivered, the course team decided that a two-course sequence was needed for the subject.

The UCLA course provides a useful case study of the objectives and topics defined for 'masters and doctoral students in information studies and in data-intensive research fields' (UCLA, 2011). Significantly, Borgman (2010) notes that development of the course drew on curricula elsewhere, but mostly created new material, having found the scope of existing courses 'did not start early enough in the scholarly process'. Her key point here is that students need to understand the lifecycle for the whole research process in order to engage effectively as a partner in data management and curation, which confirms messages from the ARL Task Force that 'science librarians…need to understand not only the concepts of the domain, but also the methodologies and norms of scholarly exchange' – a requirement

that goes well beyond the knowledge of the literature generally specified (Lougee et al., 2007, p. 5). Box 6 displays the objectives for the UCLA course.

**Course Objectives**

1. Students will learn to distinguish between the many forms of data, how data vary by scholarly discipline, and how they are used throughout the scholarly life cycle.
2. Students will learn some professional criteria for selecting and appraising data.
3. Students will learn to distinguish among different types of data collections, repositories, and services.
4. Students will learn the roles that data play in research collaborations.
5. Students will gain a basic knowledge of data curation practices in the library and archive fields.
6. Students will learn basic principles of public policies for data.

Box 6 – Course Objectives for *Data, Data Practices, and Data Curation* (UCLA, 2011)

The two-course sequence runs over two ten-week periods. Part I 'lays the foundation for data practices and services across the disciplines' and Part II 'builds upon this background to provide practical experience in data curation' (UCLA, 2011). Part I is a pre-requisite for Part II, but can also be taken on its own. Table 2 shows the topics covered in each part.

Table 2 – Topics for Parts I and II of *Data, Data Practices and Data Curation* (UCLA, 2011)

| | Part I | Part II |
|---|---|---|
| Week 1 | Overview of data, data practice, and data curation | Course introduction |
| Week 2 | Role of data in research | Selection and appraisal |
| Week 3 | [Holiday – no class] | Data collections and repositories |
| Week 4 | When are research results reproducible? | Data collections and repositories |
| Week 5 | Data lifecycle and collaboration | Technologies for data curation |
| Week 6 | Data archives and repositories | Persistent identifiers and identification |
| Week 7 | Data sharing and public policy for research data | Collection management and policy |
| Week 8 | [Holiday – no class] | Provenance and authenticity |
| Week 9 | The role of libraries and archives in data management | Formal representations |
| Week 10 | Class presentations | Student presentations |

A recurring question, posed most recently by the European report, *Riding the Wave,* (HLG, 2010) is whether our goal here is to embed data-related competencies in existing professional disciplines and programmes or, alternatively, to establish data management and curation as a new professional field – which could be seen as either a sub-field within the information profession(s) or a hybrid, 'blended professional' (Corrall, 2010), located at the intersection of two or more professional disciplines (for example library/information

science, information technology/computer science and/or archives/records management) or combining information-related expertise with an academic discipline. *Riding the Wave* asks 'How can we foster the training of more data scientists and data librarians, *as important professions in their own right?'* (HLG, 2010, p. 23, emphasis added), suggesting two new professions, as complementary career streams. However, members of the International Data curation Education Action (IDEA) Working Group have adopted a different position, expressing the opinion that 'we are training professionals to work in digital or data curation, rather than training digital or data curation professionals' (Hank and Davidson, 2009), while another perspective views data curation as an emergent specialist career *within* the relevant scientific discipline, as shown by the formation of the International Society for Biocuration in 2008 and the launch of a new specialist journal, *Database: The Journal of Biological Databases and Curation* (Howe et al., 2008).

Our discussion has concentrated on educating library and information professionals for roles in managing research data, but there is evidently scope for LIS to extend the inter-professional education that is already bringing practitioners from different information domains together (archivists, records managers, librarians, information technologists, etc) and become involved in preparing other stakeholders for their roles in managing research data. Several iSchools already offer discipline-based information programmes (e.g. Bioinformatics, Chemoinformatics, Health Informatics), but there could be opportunities for many more institutions to offer postgraduate courses and certificates to master's and doctoral students in science, social science and humanities disciplines, which would significantly benefit library and information professionals, by enabling them to gain firsthand understanding of different disciplinary perspectives on data management and curation prior to entering the field. The needs of the different cohorts would vary, but there should be sufficient common ground for some shared courses, seminars or sessions to be feasible; giving students the chance to engage in inter-professional teamwork on problems and projects would be useful preparation for supporting multi-disciplinary research in practice.

**Conclusion**

Data-intensive e-research is having a visible impact on library thinking and practice around the world. Roles and responsibilities in the management of research data are not yet settled, but such fluidity has created a real opportunity for librarians to display leadership within their institutions and the research community. Many practitioners have already demonstrated what libraries can do to manage research data in collaboration with other stakeholders. Powerful synergies exist between the longstanding library commitment to open access and the philosophy of open science, between the principles underpinning library collection management and emerging protocols for curating digital data, between the track record of libraries in technology adoption and systems development and the complex demands for integrated infrastructure and novel workflows, and between the teaching mission of librarians and the educational agenda for e-research.

Libraries have taken impressive steps to establish their place in the research data space, advancing strategically, tactically and operationally, with exemplary practices evident in areas such as policy formulation, repository development, curriculum innovation, professional updating, undergraduate learning, postgraduate training, researcher advice, data advocacy, curation profiling and resource selection. New modules, courses, programmmes, internships, institutes and fellowships have been designed and delivered for new professionals and experienced practitioners to prepare librarians and faculty for roles in data curation, digital humanities and e-science. Developments in the field have enriched our

understanding of the need for domain knowledge and technical skills in data management and flagged particular areas of competency for library attention, notably disciplinary cultures, research methodologies, scientific workflows, lifecycle management, systems thinking, technological fluency, relationship building and institutional contexts.

The debate continues around the scale of change that libraries face entering the data arena. We have confirmed that librarians have relevant experience and expertise to contribute to the data challenge. Libraries that are already involved in digital library developments, secondary data services and specialised research support are well positioned to extend their activities incrementally into managing research data. Libraries and librarians with less experience will have more to learn, but much to gain in terms of visibility, credibility and authority on campus. The whole research community is being challenged by data management and discontinuous change, but the challenge must be met and librarians can set the pace within their institution if sensibly prepared and fully committed.

**Acknowledgements**

I am indebted to Martin Lewis, Director of Library Services and University Librarian at the University of Sheffield for discussion of the role of libraries and librarians in managing research data and for use of the research data management pyramid for libraries as a framework for the central part of the chapter. I am also very grateful to Christine Borgman, Professor & Presidential Chair in Information Studies and to Jillian Wallis and Laura Wynholds, doctoral students in the Graduate School of Education and Information Studies, for their willingness to share the syllabus of the data course that they have developed at the University of California, Los Angeles.

**References**

ACRL (2000) Association of College and Research Libraries. *Information Literacy Competency Standards for Higher Education.* Chicago, IL: American Library Association, Association of College and Research Libraries. http://www.ala.org/ala/mgrps/divs/acrl/standards/informationliteracycompetency.cfm

ANDS (2011) Australian National Data Service: Governance Structure. http://ands.org.au/governance.html

ARL (2002) Association of Research Libraries. 'Collections & access for the 21st-century scholar: changing roles of research libraries, a report from the ARL Collections & Access Issues Task Force', *ARL Bimonthly Report,* 225. http://www.arl.org/bm~doc/main.pdf

Beagrie, N. (2006) 'Digital curation for science, digital libraries, and individuals', *International Journal of Digital Curation,* 1 (1), 3-16. http://www.ijdc.net/index.php/ijdc/article/viewFile/6/2

Berman, F. et al. (2002) 'The Grid: past, present, future', in Berman, F. et al. (eds) *Grid Computing – Making the Global Infrastructure a Reality*, pp. 9-49. Chichester: John Wiley & Sons.

Borgman, C. (2010) 'Why data matter to librarians – and how to educate the next generation', in Uhlir, P.F. and Levey, C.W. (eds) *The Changing Role of Libraries in Support of Research Data Activities: A Public Symposium, Meeting Recap.* Washington, DC: The National Academies, Board on Research Data and Information. http://sites.nationalacademies.org/PGA/brdi/PGA_056901

Botticelli, P. et al. (2010) 'Educating digital curators: challenges and opportunities', *6th International Digital Curation Conference, 6-8 December 2010, Chicago.*

Brophy, P. (2007) *The Library in the Twenty-First Century,* 2nd ed. London: Facet.

Canada (2011) Government of Canada. Research Data Strategy: Working Group Members. http://rds-sdr.cisti-icist.nrc-cnrc.gc.ca/eng/members.html

Carlson, J. et al. (2011) 'Determining data information literacy needs: a study of students and research faculty', *Portal: Libraries and the Academy*, 11 (2), 629–657.

Choudhury, G.S. (2008) 'Case study in data curation at Johns Hopkins University', *Library Trends,* 57 (2), 211-220.

Corrall, S. (2010) 'Educating the academic librarian as a blended professional: a review and case study', *Library Management,* 31 (8/9), 567-593.

Corrado, E.M. (2005) 'The importance of open access, open source, and open standards for libraries', *Issues in Science and Technology Librarianship,* 42. http://www.library.ucsb.edu/istl/05-spring/article2.html

Costello, K.L. and Brown, M.E. (2010) 'Preliminary report on the 2010-2011 DigCCurr Professional Institute: Curation Practices for the Digital Object Lifecycle', *D-Lib Magazine,* 16 (11/12). http://www.dlib.org/dlib/november10/costello/11costello.html

Czarnocki, S. and Khouri, A. (2004) 'A library service model for digital data support', *IASSIST* Quarterly, 28 (2), 30-34. http://www.iassistdata.org/downloads/iqvol282_3khouri.pdf

Dahlström, M. and Doracic , A. (2009) 'Digitization education: courses taken and lessons learned', *D-Lib Magazine, 1*5 (3/4). http://www.dlib.org/dlib/march09/dahlstrom/03dahlstrom.html

DCC (2010) Digital Curation Centre.*Data Management Plans.* http://www.dcc.ac.uk/resources/data-management-plans

Dempsey, L. (2003) The recombinant library: portals and people. *Journal of Library Administration,* 39 (4), 103-136.

Denniston, R., Herndon, J. and Mangiafico, P. (2011) *Research Data Planning at Duke: Subject Guide for Data Management Plans.* Durham, NC: Duke University Libraries. http://guides.library.duke.edu/dukedata

Florance, P. (2006) 'GIS collection development within an academic libray', *Library Trends,* 55 (2), 222-235.

Friedlander, A. and Adler, P. (2006) *To Stand the Test of Time: Long-term Stewardship of Digital Data Sets in Science and Engineering,* A Report to the National Science Foundation from the ARL Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe. http://www.arl.org/bm~doc/digdatarpt.pdf

Fulton, B. et al. (2011) 'DigIn: a hands-on approach to a digital curation curriculum for professional development', *Journal of Education for Library and Information Science,* 52 (2), 95-109.

Gabridge, T. (2009) 'The last mile: liaison roles in curating science and engineering research data', *Research Libraries Issues,* 265, 15-21. http://www.arl.org/bm~doc/rli-265-gabridge.pdf

Genoni, P. (2004) 'Content in institutional repositories: a collection management issue', *Library Management,* 25 (6/7), 300-306.

Gorman, M. (2000) *Our Enduring Values: Librarianship in the 21st Century.* Chicago: American Library Association.

Gray, T. (2011) *Scientific Data Repositories and Datasets.* Durham, NC: Duke University Libraries. http://guides.library.duke.edu/sciencedata

Green, A. et al. (2009) *Policy-making for Research Data in Repositories: A Guide.* Edinburgh: University of Edinburgh, Information Services Division. http://www.disc-uk.org/docs/guide.pdf

Gregory, L. and Guss, S. (2010) 'Digital curation education in practice: catching up with two former fellows', *6th International Digital Curation Conference, 6-8 December 2010, Chicago.*

Hank, C. and Davidson, J. (2009) 'International Data curation Education Action (IDEA) Working Group: a report from the second workshop of the IDEA', *D-Lib Magazine,* 15 (3/4). http://www.dlib.org/dlib/march09/hank/03hank.html

Harper, R. and Corrall, S. (2011) 'Effects of the economic downturn on academic libraries in the UK: positions and projections in mid-2009', *New Review of Academic Librarianship,* 17 (1), 96-128.

Harvey, R. (2010) *Digital Curation: A How-To-Do-It Manual.* London: Facet.

Hey, T. and Hey J**.** (2006) 'E-science and its implications for the library community', *Library Hi Tech,* 24 (4), 515-528.

Higgins, S. (2008) 'The DCC curation lifecycle model', *International Journal of Digital Curation,* 3 (1), 134-140. http://www.ijdc.net/index.php/ijdc/article/view/69

HLG (2010) High Level Expert Group on Scientific Data. *Riding the Wave: How Europe Can Gain from the Rising Tide of Scientific Data.* Brussels: European Commission. http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

Howe, D. et al. (2008) 'The future of biocuration', *Nature,* 455 (4 September), 47-50.

Hswe, P. and Holt, A. (2010) *Guide for Research Libraries: The NSF Data Sharing Policy.* Washington, DC: Association of Research Libraries. http://www.arl.org/rtl/eresearch/escien/nsf/index.shtml

Hswe, P. and Holt, A. (2011) 'Joining in the enterprise of response in the wake of the NSF data management planning requirement', *Research Library Issues,* 274, 11-17. http://publications.arl.org/17gcns.pdf

Humphrey, C. and Hamilton, E. (2004) 'is it working? Assessing the value of the Canadian Data Liberation Initiative', *The Bottom Line: Managing Library Finances,* 17 (4), 137-145.

IFLA (2011) *IFLA Statement on Open access – Clarifying IFLA's Position and Strategy.* The Hague: International Federation of Library Associations and Institutions. http://www.ifla.org/files/hq/news/documents/ifla-statement-on-open-access.pdf

Lee, H.-L. (2003) Information spaces and collections: implications for organization. *Library & Information Science Research,* 25 (4), 419-436.

Lewis, M. (2010) 'Libraries and the management of research data', in McKnight, S. (ed.), *Envisioning Future Academic Library Services: Initiatives, Ideas and Challenges,* pp. 145-168, Facet: London.

Lougee, W. et al. (2007) *Agenda for Developing E-Science in Research Libraries,* prepared by the Joint Task Force on Library Support for E-Science. Washington, DC: Association of Research Libraries. http://www.arl.org/bm~doc/ARL_EScience_final.pdf

Luce, R. (2008) 'A new value equation challenge: the emergence of eResearch and roles for research libraries', in Council on Library and Information Resources, *No Brief Candle: Reconceiving Research Libraries for the 21st Century,* pp. 42-50. Washington, DC: CLIR. http://www.clir.org/pubs/reports/pub142/pub142.pdf

Lynch, C. (2003) Institutional repositories: essential infrastructure for scholarship in the digital age. *ARL Bimonthly Report,* 226, 1-7. http://www.arl.org/bm~doc/br226ir.pdf

Lyon, L. (2007) *Dealing with Data: Roles, Rights, Responsibilities and Relationships,* Consultancy Report, Bath: UKOLN. http://www.ukoln.ac.uk/ukoln/staff/e.j.lyon/reports/dealing_with_data_report-final.pdf

Macdonald, S. and Martinez, L. (2005) 'Supporting local data users in the UK academic community', *Ariadne,* 44. http://www.ariadne.ac.uk/issue44/martinez/

MacDonald, S. and Uribe, L.M. (2008) 'Libraries in the converging world of open data, e-research, and Web 2.0', *Online,* 32 (2), 36-40. http://ie-repository.jisc.ac.uk/227/1/Online_mar08.pdf

MacMillan, D. (2010) 'Sequencing genetics information: integrating data into information literacy for undergraduate biology students', *Issues in Science and Technology Librarianship,* 61 (Spring). http://www.istl.org/10-spring/refereed3.html

Martinez, L. (2007) *The e-Research Needs Analysis Survey Report.* CURL/SCONUL Joint Task Force on e-Research. http://www.rluk.ac.uk/files/E-ResearchNeedsAnalysisRevised.pdf

Martinez-Uribe, L. (2007) 'Digital repository services for managing research data: what do Oxford researchers need?', *IASSIST Quarterly,* 31 (3-4), 28-33. http://www.iassistdata.org/publications/iq/iq31/iqvol313martinez.pdf

MIT (2011) Data Management Plans. Cambridge, MA: Massachusetts Institute of Technology Libraries. http://libraries.mit.edu/guides/subjects/data-management/plans.html

Newton, M.P. et al. (2010) 'Librarian roles in institutional repository data set collecting: outcomes of a research library task force', *Collection Management,* 36 (1), 53-67.

Nicholas, D. et al. (2010) 'The impact of the economic downturn on libraries: with special reference to university libraries', *Journal of Academic Librarianship,* 36 (5), 376-382.

NSB (2005) National Science Board. *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century.* Arlington, VA: National Science Foundation, National Science Board. http://www.nsf.gov/pubs/2005/nsb0540.pdf

NSF (2007) National Science Foundation. *Cyberinfrastructure Vision for 21st Century Discovery.* Arlington, VA: National Science Foundation, Cyberinfrastructure Council. http://www.nsf.gov/pubs/2007/nsf0728/nsf0728.pdf

Ogburn, J.L. (2010) 'The imperative for data curation', *Portal: Libraries and the Academy,* 10 (2), 241-246.

Pryor, G. and Donnelly, M. (2009), 'Skilling up to do data: whose role, whose responsibility, whose career?', *International Journal of Digital Curation,* 4 (2), 158-170. http://www.ijdc.net/index.php/ijdc/article/view/126

Purdue University Libraries (2010) *Data Curation Profiles Toolkit.* West Lafayette, IN: Purdue University Libraries. http://www4.lib.purdue.edu/dcp/

Purdue University Libraries (2011) Data Management Plan Self-Assessment Questionnaire. West Lafayette, IN: Purdue University, Purdue University Libraries/Office of the Vice President for Research. http://d2c2.lib.purdue.edu/documents/dmp_self_questionnaire.pdf

Qin, J. et al (2010) 'Educating eScience Librarians', Poster, *6th International Digital Curation Conference, 6-8 December 2010, Chicago.* http://meganoakleaf.info/educatingsciencelibs.pdf

Ray, J. (2009) 'Sharks, digital curation, and the education of information professionals', *Museum Management and Curatorship,* 24 (4), 357-368.

Read, E.J. (2007) 'Data services in academic libraries: assessing needs and promoting services', *Reference & User Services Quarterly,* 46 (3), 61-75.

Renear, A.H. et al. (2010) 'Extending an LIS data curation curriculum to the humanities: selected activities and observations', Poster Abstract, *iConference 2010 Proceedings, February 3-6, University of Illinois at Urbana Champaign.* https://www.ideals.illinois.edu/handle/2142/15061

Research Information Network (2008) *Mind the Skills Gap: Information-Handling Training for Researchers.* London: Research Information Network. www.rin.ac.uk/system/files/attachments/**Mind-skills-gap**-report.pdf

Rice, R. (2009) *DISC-UK DataShare Project: Final Report.* Joint Information Systems Committee. http://ie-repository.jisc.ac.uk/336/1/DataSharefinalreport.pdf

Rieh, S.Y. et al. (2007) Census of institutional repositories in the U.S.: a comparison across institutions at different stages of IR development, *D-Lib Magazine,* 13 (11/12). http://www.dlib.org/dlib/november07/rieh/11rieh.html

Salo, D. (2010) ' Retooling libraries for the data challenge' *Ariadne,* 64. http://www.ariadne.ac.uk/issue64/salo/

SCONUL (2011) *The SCONUL Seven Pillars of Information Literacy: A Research Lens for Higher Education.* London: Society of College, National and University Libraries, Working Group on Information Literacy. http://www.sconul.ac.uk/groups/information_literacy/publications/researchlens.pdf

Shearer, K. and Argáez, D. (2010) *Addressing the Research Data Gap: A Review of Novel Services for Libraries.* Ottawa: Canadian Association of Research Libraries, http://www.carl-abrc.ca/about/working_groups/pdf/library_roles-final.pdf

Soehner, C. et al. (2010) *E-Science and Data Support Services: A Study of ARL Member Institutions.* Washington, DC: Association of Research Libraries. http://www.arl.org/bm~doc/escience_report2010.pdf

Stanton, J.M. et al. (2011) 'Education for eScience professionals: job analysis, curriculum guidance, and program considerations', *Journal of Education for Library and Information Science,* 52 (2), 79-94.

Stephenson, E. and Caravello, P.S. (2007) 'Incorporating data literacy into undergraduate information literacy programs in the social sciences: a pilot project', *Reference Services Review*, 35 (4), 525-540.

Swan, A. and Brown, S. (2008) *The Skills, Role and Career Structure of Data Scientists and Curators: An Assessment of Current Practice and Future Needs.* Truro: Key Perspectives. http://www.jisc.ac.uk/publications/documents/dataskillscareersfinalreport.aspx

Syracuse University (2010) 'Syracuse iSchool and Cornell University launch eScience mentorship program'. Syracuse, NY: Syracuse University School of Information Studies. http://ischool.syr.edu/newsroom/?recid=980

Szigeti, K. and Wheeler, K. (2011) 'Essential readings in e-science', *Issues in Science and Technology Librarianship,* 64 (Winter). http://www.istl.org/11-winter/internet2.html

Takeda, K. (2010) *Institutional Data Management Blueprint: Initial Findings Report.* Southampton: University of Southampton. http://www.southamptondata.org/1/post/2010/12/idmb-initial-findings-report.html

UCLA (2011) University of California, Los Angeles. *Data, Data Practices, and Data Curation. Parts I and II.* Unpublished document.

University of Connecticut (2011) Data Management Plans for Grant Funded Projects (NSF, NIH). Storrs, CT: University of Connecticut Libraries. http://www.lib.uconn.edu/scholarlycommunication/data.html

University of Edinburgh (2011) Policy for Management of Research Data. Edinburgh: The University. http://www.docs.is.ed.ac.uk/docs/data-library/rdm-policy.pdf

University of Massachusetts (2011) eScience Portal for New England Librarians: A Librarian's Link to eScience Resources. Worcester, MA: University of Massachusetts Medical School. http://escience library.umassmed.edu/index

UW-Madison (2011a) Research Data Services: About Us. Madison, WI: University of Wisconsin-Madison. http://researchdata.wisc.edu/get-help/about-us/

UW-Madison (2011b) Research Data Services: Our Services. Madison, WI: University of Wisconsin-Madison. http://researchdata.wisc.edu/get-help/our-services/

Varalakshmi, R.S.R. (2009) 'Curriculum for digital libraries: an analytical study of Indian LIS curricula', *D-Lib Magazine,* 15 (9/10). http://www.dlib.org/dlib/september09/varalakshmi/09varalakshmi.html

Vitae (2010) *Researcher Development Framework.* Cambridge: Vitae/Careers Research and Advisory Centre. http://www.vitae.ac.uk/rdf

Walters, T. and Skinner, K. (2011) *New Roles for New Times: Digital Curation for Preservation.* Washington, DC: Association of Research Libraries. http://www.arl.org/bm~doc/nrnt_digital_curation17mar11.pdf

Walters, T.O. (2009) 'Data curation program development in U.S. universities: the Georgia Institute of Technology example', *International Journal of Digital Curation,* 3 (4), 83-92. http://www.ijdc.net/index.php/ijdc/article/viewFile/136/153].

Westra, B. (2010) 'Data services for the sciences: a needs assessment.' *Ariadne,* 64. http://www.ariadne.ac.uk/issue64/westra/

Westra, B. et al. (2010) 'Selected Internet resources on digital research data curation', *Issues in Science and Technology Librarianship,* 63 (Fall). http://www.istl.org/10-fall/internet2.html

Whyte, A. and Wilson, A. (2010) *How to Appraise & Select Research Data for Curation.* Digital Curation Centre and Australian National Data Service. Available at: http://www.dcc.ac.uk/node/9098#c9

Willinsky, J. (2005) 'The unacknowledged convergence of open source, open access, and open science', *First Monday,* 10 (8). http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1265

Witt, M. (2008) 'Institutional repositories and research data curation in a distributed environment', *Library Trends,* 57 (2), 191-201.

Witt, M. and Carlson, J.R. (2007) 'Conducting a data interview', Poster/Libraries Research Publications. West Lafayette, IN: Purdue University Libraries. http://docs.lib.purdue.edu/lib_research/81/

Yakel, E. et al. (2011) 'Digital curation for digital natives', *Journal of Education for Library and Information Science,* 52 (1), 23-31.